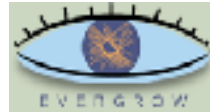


Evergrow Workshop

Jerusalem June 2006



Numerical Applications

@

Roma I “La Sapienza”, ISI Foundation Turin, ICTP Trieste

Andrea Pagnani

ISI Foundation Turin

OUTLINE

- Exact computation of 2D Spin Glasses Rome
- Optimized Monte Carlo methods Rome
- Exact enumeration of circuits in graphs Rome
- Statistical properties of d-dimensional KPZ surfaces Rome, Turin
- Inference of Gene Regulatory Networks from gene expression patterns Trieste, Turin

Optimized Monte Carlo methods

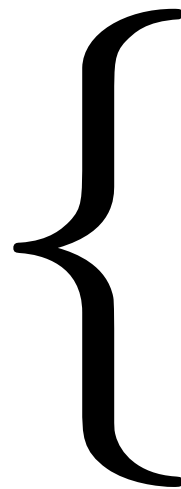
- Use of parallel tempering and replica cluster Monte Carlo strategy for determining the low temperature properties of finite dimensional Spin Glass models
- Critical behavior of 3D Edwards-Anderson model and universality of the finite-size scaling functions
- Check of universality: Do critical exponents depend on the distribution of the couplings? Site-diluted gaussian vs. bond-diluted

Exact computation of 2D Spin Glasses

The critical behavior of the 2D Edwards-Anderson model

$T=0$ (optimal solutions) vs. $T>0$ (good solutions)

2 different strategies



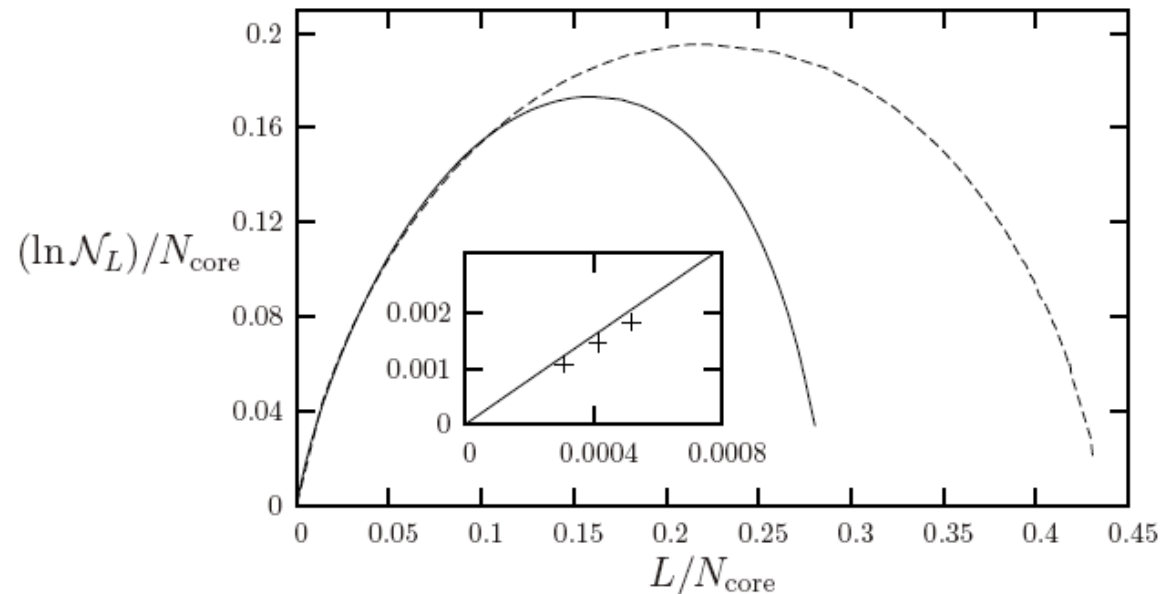
Replica Cluster Method
[heuristic]

Partition function solvers
matching problem
[exact]

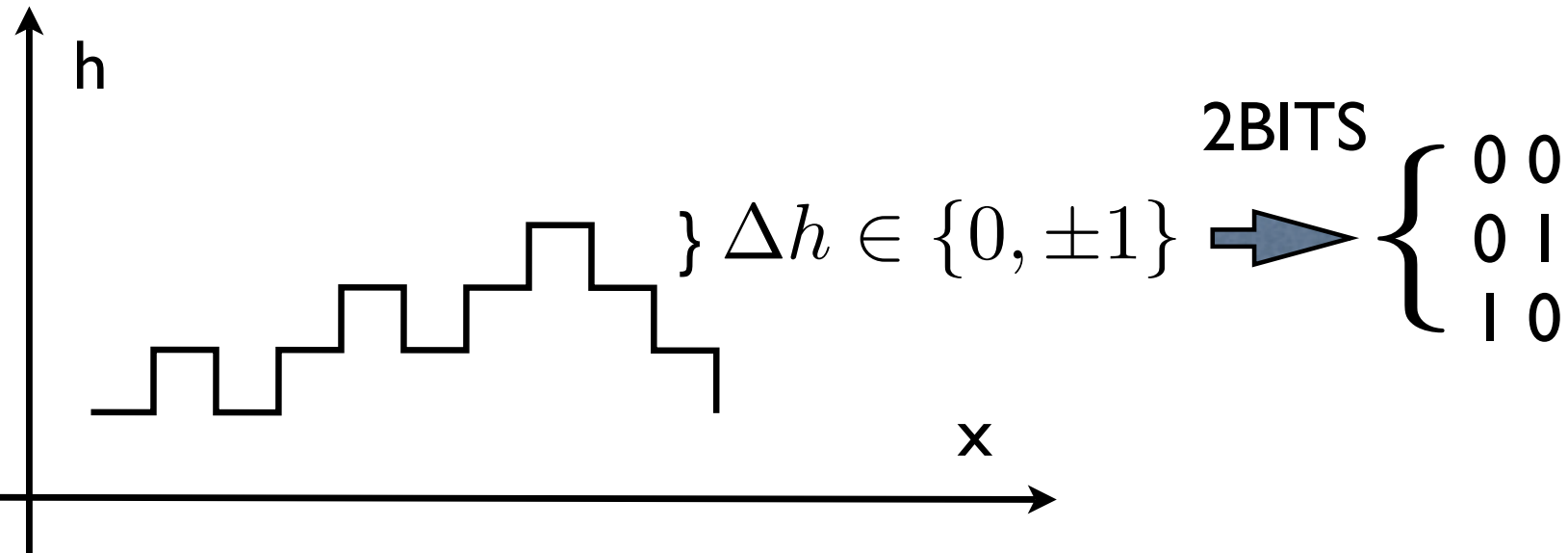
Geometrical organization of low
temperature states

Circuit enumeration in graphs

- Exact enumeration: feasible on large graphs only for loops of size $\sim 3,4,5$
- Message-Passing algorithms estimate the number of circuits of any size (DIMES data)
- Try to build an efficient Monte Carlo method for loop counting.
- Decimation strategy allows for singling out loops of a given length



Statistical properties of d-dimensional KPZ surfaces

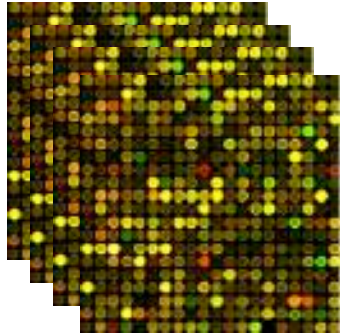


Strategy: 1 int \rightarrow 32 or 64 surface points

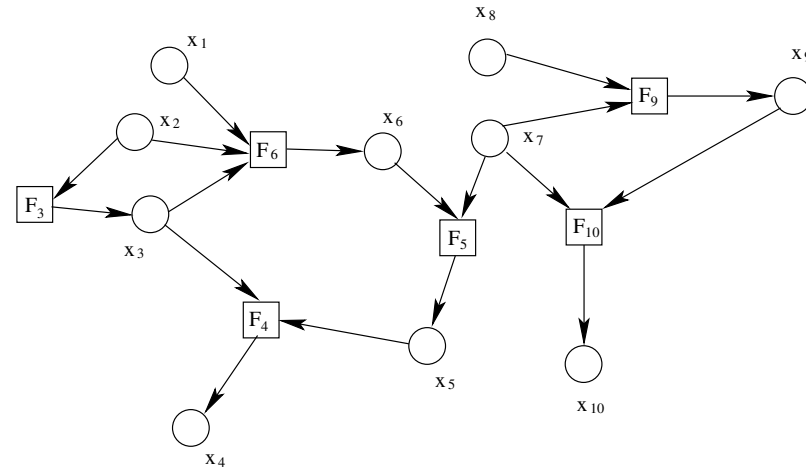
From the moments w_k we can estimate the critical exponent χ $\left\{ \begin{array}{l} w_k(t) = \frac{1}{V} \sum_{i=1}^V [(h_i(i) - \bar{h}_i(t))^k] \\ \lim_{t \rightarrow \infty} w_k(t) = A_k L^{k\chi} (1 + B_k L^{-\omega}) \end{array} \right.$

We can simulate lattice of size $L=192^4$

Statistical Inference of Gene Regulatory Networks



M gene profiles



M expression profiles of N+1 genes $(x_0^a, \dots, x_{N+1}^a)$, $a \in (1, \dots, N)$

Which genes regulate gene x_0 ? \longrightarrow Network Topology

How genes regulate x_0 ? \longrightarrow Combinatorial Control

$$H[\mathbf{x}_0, \mathbf{J}] = \sum_{a=1}^M \left[x_0^a - \sum_{i=1}^N (f^+(x_i^a) \delta_{J_i, +1} + f^-(x_i^a) \delta_{J_i, -1}) \right]^2 \quad J \in \{0, \pm 1\}$$

Realistic value (B-Cell in human) N=12000 M=350

In conclusion: what do we need?

- A fast numerical factory. We need to run many samples (quenched averages), embarrassing parallelism.
- Shell scripting is almost always widely used.
- Limited I/O. Important memory usage.
- Load balancing?
- Job scheduling?
- What about the use of MPI (real parallelism)?